

Big Data

- 1 A lot of data
- 2 data Complexity
- 3 Variety of data
- 4 data Streaming

1 Structured  
4 Unstructured # XML  
2 Semistructured  
3 Quasi Structured

web search history  
web click stream

test

3 V model

Volume Vareity Velocity

Value of data  $\equiv$  Business value

4 V model

Volume Vareity Velocity Value

Certainty of data  $\equiv$  Veracity

5 V model

Volume Vareity Velocity Value Veracity

R lang.  
 $\equiv$   
Tool

test  
ch/c of Big data

Module 1

distribution of data  $\rightarrow$

time lines  $\rightarrow$  rate of data creation, every day/month/year/sec

data Analy scientist  $\rightarrow$  must be doubtful  
 $\rightarrow$  must be aware of lots of fields

OLAP  $\rightarrow$  Online Analytical Processing

test  
Slide 17

Data island  $\rightarrow$  not connected, easy, any one can make it  
not secured, replication may exist  
(redundancy), data can get lost

data warehouse  $\rightarrow$  secure, Backup exists, connected  
DB Administrator, is in Control, No replication  
 $\rightarrow$  responsible for permissions



Sandbox → Analyst owns it not Administrator  
 Big Size, not small  
 Variety of data, more accurate results.  
 Shadow file sys. → I see the data in the Data warehouse but don't have a replication of it & can't change it without permission

### mini-Case Study requirements

- more security
- relations between ~~connected~~ currencies
- laws / legal requirements
- distributed storage
- Controlled access to data from all Banks
- more services (ATM) (Employees)

requirements from Analyst point of view:

- Velocity
- Variety (اكثر تنوع)
- Volume
- Tools & methods to manage the data
- data driven (manages based on the data he gets)

KPI ≡ Key Performance Indicator

TP True +ve  
 FN False -ve  
 TN True -ve  
 FP False +ve

$$\frac{TP + TN}{\text{total}} = \text{accuracy}$$

Business intelligence, data Analyst <sup>oral best</sup> ~~Il~~ <sup>من</sup> ~~نرى~~ <sup>نرى</sup>

- Skills
- how can they help you

~~Business~~

Business Intelligence → Just an indicator  
 doesn't take decisions  
 uses KPI, deals with structured data

Data Analyst → ~~gives~~ Analysis &  
 gives decisions

High knowledge in AI, machine learning, Math, data mining